Speeda: Adaptive Speed-up for Lecture Videos

Chen-Tai Kao, Yen-Ting Liu, Alexander Hsu
Computer Science Department
Stanford University
{chentai, eggegg, kzm}@stanford.edu

ABSTRACT
Increasing the playback speed of lecture videos is a common technique to shorten watching time. This creates challenges when part of the lecture becomes too fast to be discernible, even if the overall playback speed is acceptable. In this paper, we present a speed-up system that preserves lecture clearness in high playback rate. A user test was conducted to evaluate the system. The result indicates that our system significantly improves user's comprehension level.

Author Keywords
MOOC; playback; speed

ACM Classification Keywords
H.5.2. User Interfaces: User-centered design

INTRODUCTION
Increasing popularity of Massive Open Online Course (MOOC) has resulted in a variety of user behaviors intended to boost study efficiency. Using higher playback speed is a common technique to save precious watching time, as long as the overall lecture remains understandable. However, due to fluctuating speaking rate, sometimes it becomes impossible to understand video segments where the speaker talks fast. Also, there exists room to accelerate video segments with little information (e.g. long pauses) to further save time. These problems inevitably lower both study efficiency and quality.

Several works focused on video and audio summarization [1] [5], using techniques including time compression and pause reduction. However, crucial information may be missing in video summaries, making the lecture inconsistent. SmartPlayer [2] achieved adaptive video fast-forward based on the complexity of current scene and predefined semantic events. In this manner, users were able to scan through video faster without missing areas of interest. While the work considered video complexity, it did not exploit speech features, which rendered the lecture impossible to listen.

In this paper, we propose that the playback speed should be dynamically adjusted based on instantaneous syllable density.

UIST '14, Oct 05-08 2014, Honolulu, HI, USA
ACM 978-1-4503-3068-8/14/10.
http://dx.doi.org/10.1145/2658779.2658794

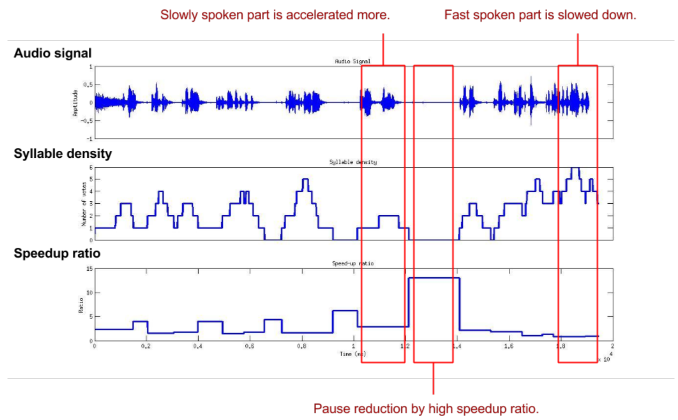

Figure 1. An audio sample processed by Speeda. Top to bottom: The original audio signal; instantaneous syllable density; speed-up ratio of each clip. Note that the speed-up ratio is calculated inversely proportional to the speaking rate, which is indicated by syllable density.

That is, video segments with few syllables should be accelerated more, and vice versa. Consequently, playback time can be minimized while the lecture remains clear enough to understand.

We introduce Speeda, a video speed-up system that implements this concept. With wide adoption of MOOC videos, Speeda is able to save large amount of learning time and boost learning quality simultaneously.

SYSTEM DESIGN AND IMPLEMENTATION
To ensure every part of the video is discernible, the speed-up ratio is made inversely proportional to the syllable density. See Figure 2 for architecture of Speeda, which we explain below.

First, syllables detected by Harma's method [4] are used for segmentation, generating lots of *clips*. There are two kinds of *clip*, where a *segment* contains several syllables, and a *pause* has none. For every *segment*, the speed-up ratio is determined by its average syllable density as follows:

$$\text{speedup ratio} = \text{desired speed} \times \frac{\text{segment's syllable density}}{\text{video's syllable density}}.$$

All *pauses* are sped up to have 0.2 seconds, which is short but good enough for breaking in sentences. Now the speed-up ratio of all segments and pauses are determined, we speed

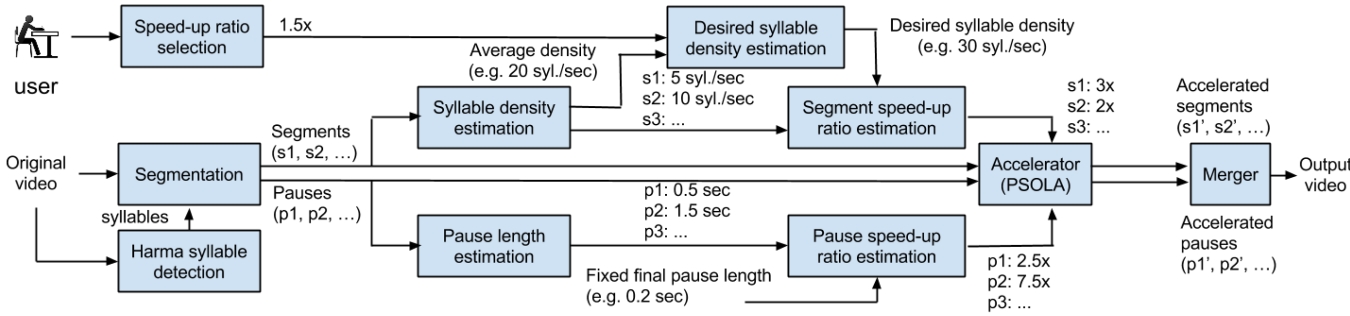

Figure 2. The Speeda system. Given the desired speed, Speeda calculates each clip's speed-up ratio and accelerates the video accordingly.

up each clip accordingly by Pitch Synchronous Overlap and Add (PSOLA) [3], a time stretching technique that changes speed of an audio signal without affecting its pitch. We then concatenate all the clips to generate the output video. Figure 1 gives an example of a 10-second speech sample processed by Speeda.

EVALUATION AND RESULT

Our research question was whether video generated by Speeda improves user's comprehension level compared to normal speed-up video of the same length. To answer the question, we picked a TED talk video no participant has watched before. The talk was about ethics, reducing the bias of participant's prior knowledge. Both the standard 2x speed-up version and an Speeda version were generated and had the same length. Lasecki et al. [6] used comprehension test to evaluate the impact of various video caption systems on students' comprehension level. Similar method was adopted here to evaluate Speeda.

10 students participated our user study (5 females, age ranged from 19 to 35). All of them had experience watching MOOC videos, spending 1-3 hours per week in average. Participants were randomly divided into two groups. We conducted a between-subject study, where one group watched normal speed-up video, and the other watched the Speeda version. After that, we asked 6 comprehension questions, with score ranging from 0-6, which was the number of questions correctly answered.

We found comprehension improvement for users who watched the Speeda version. A t-test comparing the comprehension score of the control (normal speed-up, mean = 2, std. dev = 0) to the scores with the treatment condition (Speeda speed-up, mean = 3.6, std. dev = 1.1402) is significant (t(-3.1379) = 0.03492, p¡.05).

DISCUSSION

The study result indicates that Speeda improves comprehension. Surprisingly, our method is not overwhelmingly favored by all participants. Some reported that the video sped up by Speeda sounds slightly unnatural. How to improve the video quality requires further exploration.

Although the video content accelerated by Speeda were better understood, note-taking under such high playback speed was still stressful, as suggested by some feedback. This can be expected, given that the lecture time is drastically compressed. It thus highlights the potential to explore how to pause lectures adaptive to its context.

CONCLUSION

In this paper we presented a speed-up system that plays lecture fast and clear, by using syllable density to dynamically determine the speed-up ratio. We believe that the proposed system has several strengths. First, fast spoken parts are slowed down, greatly boosting learning quality. Second, slowly spoken parts are accelerated even more, saving precious watching time. Lastly, as shown by a user study, Speeda significantly enhanced user's comprehension level.

REFERENCES

1. Arons, B. Speechskimmer: A system for interactively skimming recorded speech. *ACM Trans. Comput.-Hum. Interact. 4*, 1 (Mar. 1997), 3–38.

2. Cheng, K.-Y., Luo, S.-J., Chen, B.-Y., and Chu, H.-H. Smartplayer: User-centric video fast-forwarding. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '09, ACM (New York, NY, USA, 2009), 789–798.

3. Gold, B., Morgan, N., and Ellis, D. *Speech and Audio Signal Processing: Processing and Perception of Speech and Music*, 2nd ed. Wiley-Interscience, New York, NY, USA, 2011.

4. Harma, A. Automatic identification of bird species based on sinusoidal modeling of syllables. In *Acoustics, Speech, and Signal Processing, 2003. Proceedings. (ICASSP '03). 2003 IEEE International Conference on*, vol. 5 (April 2003), V–545–8 vol.5.

5. He, L., Sanocki, E., Gupta, A., and Grudin, J. Auto-summarization of audio-video presentations. In *Proceedings of the Seventh ACM International Conference on Multimedia (Part 1)*, MULTIMEDIA '99, ACM (New York, NY, USA, 1999), 489–498.

6. Lasecki, W. S., Kushalnagar, R., and Bigham, J. P. Helping students keep up with real-time captions by pausing and highlighting. In *Proceedings of the 11th Web for All Conference*, W4A '14, ACM (New York, NY, USA, 2014), 39:1–39:8.